OPPA European Social Fund
Prague & EU: We invest in your future.

Primer on Probability for Discrete Variables

BMI/CS 576
www.biostat.wisc.edu/bmi576.html
Mark Craven
craven@biostat.wisc.edu
Fall 2011

Definition of probability

- *frequentist* interpretation: the probability of an event from a random experiment is the proportion of the time events of same kind will occur in the long run, when the experiment is repeated
- examples
  - the probability my flight to Chicago will be on time
  - the probability this ticket will win the lottery
  - the probability it will rain tomorrow
- always a number in the interval [0,1]
  0 means "never occurs"
  1 means "always occurs"

Sample spaces

- *sample space*: a set of possible outcomes for some event

- examples
  - flight to Chicago: {on time, late}
  - lottery: {ticket 1 wins, ticket 2 wins,…,ticket *n* wins}
  - weather tomorrow:

    {rain, not rain} or

    {sun, rain, snow} or

    {sun, clouds, rain, snow, sleet} or…

Random variables

- *random variable*: a variable representing the outcome of an experiment

- example
  - X represents the outcome of my flight to Chicago
  - we write the probability of my flight being on time as $P(X = \text{on-time})$
  - or when it's clear which variable we're referring to, we may use the shorthand $P(\text{on-time})$

Notation

- uppercase letters and capitalized words denote random variables
- lowercase letters and uncapitalized words denote values
- we'll denote a particular value for a variable as follows

$$P(X = x) \qquad P(Fever = true)$$

- we'll also use the shorthand form

$$P(x) \quad \text{for} \quad P(X = x)$$

- for Boolean random variables, we'll use the shorthand

$$P(fever) \quad \text{for} \quad P(Fever = true)$$

$$P(\neg fever) \quad \text{for} \quad P(Fever = false)$$

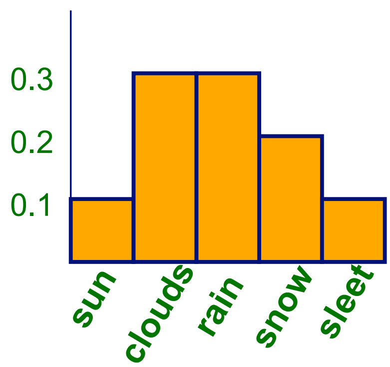
Probability distributions

- if X is a random variable, the function given by $P(X = x)$ for each x is the *probability distribution* of X

- requirements:

$$P(x) \geq 0 \quad \text{for every } x$$

$$\sum_x P(x) = 1$$



Joint distributions

- *joint probability distribution*: the function given by $P(X = x, Y = y)$
- read "X equals x <u>and</u> Y equals y"
- example

x, y	$P(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

probability that it's sunny and my flight is on time (→ 0.20)

Marginal distributions

- the *marginal distribution* of X is defined by

$$P(x) = \sum_y P(x, y)$$

"the distribution of X ignoring other variables"

- this definition generalizes to more than two variables, e.g.

$$P(x) = \sum_y \sum_z P(x, y, z)$$

Marginal distribution example

joint distribution

x, y	$P(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distribution for X

x	$P(X = x)$
sun	0.3
rain	0.5
snow	0.2

Conditional distributions

- the *conditional distribution* of X given Y is defined as:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

"the distribution of X given that we know the value of Y"

Conditional distribution example

joint distribution

x, y	$P(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

conditional distribution for X given Y=on-time

x	$P(X = x \mid Y = on\text{-}time)$
sun	0.20/0.45 = 0.444
rain	0.20/0.45 = 0.444
snow	0.05/0.45 = 0.111

Independence

- two random variables, X and Y, are *independent* if

$$P(x,y) = P(x) \times P(y) \quad \text{for all } x \text{ and } y$$

Independence example #1

joint distribution

x, y	$P(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distributions

x	$P(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$P(Y = y)$
on-time	0.45
late	0.55

Are X and Y independent here? NO.

Independence example #2

joint distribution

x, y	$P(X = x, Y = y)$
sun, fly-United	0.27
rain, fly-United	0.45
snow, fly-United	0.18
sun, fly-Northwest	0.03
rain, fly-Northwest	0.05
snow, fly-Northwest	0.02

marginal distributions

x	$P(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$P(Y = y)$
fly-United	0.9
fly-Northwest	0.1

Are X and Y independent here? YES.

Conditional independence

- two random variables X and Y are *conditionally independent* given Z if

$$P(X \mid Y, Z) = P(X \mid Z)$$

"once you know the value of Z, knowing Y doesn't tell you anything about X"

- alternatively

$$P(x, y \mid z) = P(x \mid z) \times P(y \mid z) \quad \text{for all } x, y, z$$

Conditional independence example

Flu	Fever	Vomit	P
true	true	true	0.04
true	true	false	0.04
true	false	true	0.01
true	false	false	0.01
false	true	true	0.009
false	true	false	0.081
false	false	true	0.081
false	false	false	0.729

Are Fever and Vomit independent? NO.

e.g. $P(fever, vomit) \neq P(fever) \times P(vomit)$

Conditional independence example

Flu	Fever	Vomit	*P*
true	true	true	0.04
true	true	false	0.04
true	false	true	0.01
true	false	false	0.01
false	true	true	0.009
false	true	false	0.081
false	false	true	0.081
false	false	false	0.729

Are Fever and Vomit conditionally independent given Flu: YES.

$$P(fever, vomit \mid flu) = P(fever \mid flu) \times P(vomit \mid flu)$$
$$P(fever, vomit \mid \neg flu) = P(fever \mid \neg flu) \times P(vomit \mid \neg flu)$$

etc.

Chain rule of probability

- for two variables

$$P(X,Y) = P(X \mid Y) \times P(Y)$$

- for three variables

$$P(X,Y,Z) = P(X \mid Y,Z) \times P(Y \mid Z) \times P(Z)$$

- etc.
- to see that this is true, note that

$$P(X,Y,Z) = \frac{P(X,Y,Z)}{P(Y,Z)} \times \frac{P(Y,Z)}{P(Z)} \times P(Z)$$

Bayes theorem

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)} = \frac{P(y \mid x)P(x)}{\sum_x P(y \mid x)P(x)}$$

- this theorem is extremely useful
- there are many cases when it is hard to estimate $P(x \mid y)$ directly, but it's not too hard to estimate $P(y \mid x)$ and $P(x)$

Bayes theorem example

- MDs usually aren't good at estimating $P(Disorder \mid Symptom)$
- they're usually better at estimating $P(Symptom \mid Disorder)$
- if we can estimate $P(Fever \mid Flu)$ and $P(Flu)$ we can use Bayes' Theorem to do diagnosis

$$P(flu \mid fever) = \frac{P(fever \mid flu)P(flu)}{P(fever \mid flu)P(flu) + P(fever \mid \neg flu)P(\neg flu)}$$

Expected values

- the *expected value* of a random variable that takes on numerical values is defined as:

$$E[X] = \sum_x x \times P(x)$$

this is the same thing as the *mean*

- we can also talk about the expected value of a function of a random variable

$$E[g(X)] = \sum_x g(x) \times P(x)$$

Expected value examples

$$E[Shoesize] =$$
$$5 \times P(Shoesize = 5) + \ldots + 14 \times P(Shoesize = 14)$$

- Suppose each lottery ticket costs $1 and the winning ticket pays out $100. The probability that a particular ticket is the winning ticket is 0.001.

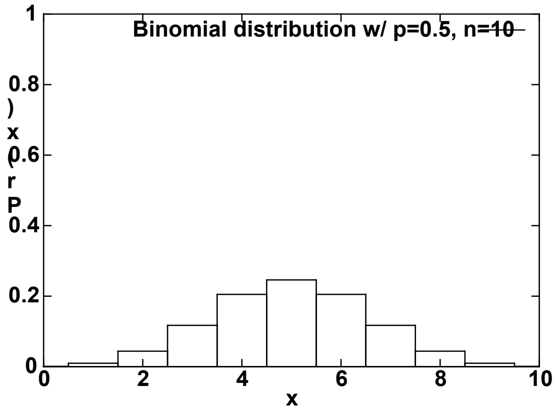
$$E[gain(Lottery)] =$$
$$gain(\text{winning})P(\text{winning}) + gain(\text{losing})P(\text{losing}) =$$
$$(\$100 - \$1) \times 0.001 - \$1 \times 0.999 =$$
$$-\$0.90$$

The binomial distribution

- distribution over the number of successes in a fixed number n of independent trials (with same probability of success p in each)

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- e.g. the probability of x heads in n coin flips

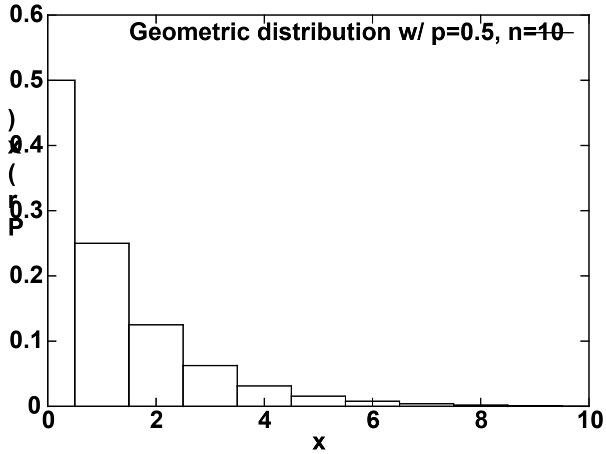


The geometric distribution

- distribution over the number of trials before the first failure (with same probability of success p in each)

$$P(x) = (1-p)p^x$$

- e.g. the probability of x heads before the first tail



The multinomial distribution

- k possible outcomes on each trial
- probability p_i for outcome x_i in each trial
- distribution over the number of occurrences x_i for each outcome in a fixed number n of independent trials

$$P(\mathbf{x}) = \frac{n!}{\prod_i (x_i!)} \prod_i p_i^{x_i}$$

vector of outcome occurrences

- e.g. with k=6 (a six-sided die) and n=30

$$P([7,3,0,8,10,2]) = \frac{30!}{7!\times 3!\times 0!\times 8!\times 10!\times 2!}\left(p_1^{\,7} p_2^{\,3} p_3^{\,0} p_4^{\,8} p_5^{\,10} p_6^{\,2}\right)$$

OPPA European Social Fund
Prague & EU: We invest in your future.